

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/409,122

DATE: 01/29/97

TIME: 13:48:15

INPUT SET: S15202.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages.

SEQUENCE LISTING

(1) General Information

(i) APPLICANT: JOYCE, JAMES G.
GEORGE, HUGH A.
HOFMANN, KATHRYN J.
JANSEN, KATHRIN U.
NEEPER, MICHAEL P.

Does Not Comply
Corrected Diskette Needed

(ii) TITLE OF THE INVENTION: DNA ENCODING HUMAN PAPILLOMAVIRUS TYPE 18 VACCI

(iii) NUMBER OF SEQUENCES: 16

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: CHRISTINE E. CARTY - MERCK & CO., INC.
(B) STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
(C) CITY: RAHWAY
(D) STATE: NJ
(E) COUNTRY: US
(F) ZIP: 07065-0907

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 08/408,669
(B) FILING DATE: 22-MAR-1995
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

Prior

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: CARTY, CHRISTINE E
(B) REGISTRATION NUMBER: 36,099
(C) REFERENCE/DOCKET NUMBER: 19425

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 908-594-6734

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/409,122

DATE: 01/29/97

TIME: 13:48:17

INPUT SET: S15202.raw

47 (B) TELEFAX: 908-594-4720
48 (C) TELEX:
49
50

51 (2) INFORMATION FOR SEQ ID NO:1:
52

53 (i) SEQUENCE CHARACTERISTICS:

54 (A) LENGTH: 1524 base pairs

55 (B) TYPE: nucleic acid

56 (C) STRANDEDNESS: single

57 (D) TOPOLOGY: linear
58

59 (ii) MOLECULE TYPE: cDNA

60 (iii) HYPOTHETICAL: NO

61 (iv) ANTI-SENSE: NO

62 (v) FRAGMENT TYPE:

63 (vi) ORIGINAL SOURCE:
64

65 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
66

67	ATGGCTTTGT	GGCGGCCTAG	TGACAATACC	GTATACCTTC	CACCTCCTTC	TGTGGCAAGA	60
68	GTTGTAAATA	CTGATGATTA	TGTGACTCGC	ACAAGCATAT	TTTATCATGC	TGGCAGCTCT	120
69	AGATTATTAA	CTGTTGGTAA	TCCATATTTT	AGGGTTCCTG	CAGGTGGTGG	CAATAAGCAG	180
70	GATATTCCTA	AGGTTTCTGC	ATACCAATAT	AGAGTATTTT	GGGTGCAGTT	ACCTGACCCA	240
71	AATAAATTTG	GTTTACCTGA	TAATAGTATT	TATAATCCTG	AAACACAACG	TTTAGTGTGG	300
72	GCCTGTGCTG	GAGTGGAAAT	TGGCCGTGGT	CAGCCTTTAG	GTGTTGGCCT	TAGTGGGCAT	360
73	CCATTTTATA	ATAAATTAGA	TGACACTGAA	AGTTCCCATG	CCGCTACGTC	TAATGTTTCT	420
74	GAGGACGTTA	GGGACAATGT	GTCTGTAGAT	TATAAGCAGA	CACAGTTATG	TATTTTGGGC	480
75	TGTGCCCCCTG	CTATTGGGGA	ACACTGGGCT	AAAGGCACTG	CTTGTAATC	GCCTCCTTTA	540
76	TCACAGGGCG	ATTGCCCCCC	TTTAGAACTT	AAGAACACAG	TTTTGGGAAGA	TGGTGATATG	600
77	GTAGATACTG	GATATGGTGC	CATGGACTTT	AGTACATTGC	AAGATACTAA	ATGTGAGGTA	660
78	CCATTGGATA	TTTGTCACTC	TATTTGTAAA	TATCCTGATT	ATTTACAAAT	GTCTGCAGAT	720
79	CCTTATGGGG	ATTCATGTTT	TTTTTGCCTT	CGACGTGAGC	AGCTTTTTGC	TAGGCATTTT	780
80	TGGAATAGGG	CAGGTACTAT	GGGTGACACT	GTGCCTCAAT	CCTTATATAT	TAAAGGCACA	840
81	GGTATGCGTG	CTTCACCTGG	CAGCTGTGTG	TATTCTCCCT	CTCCAAGTGG	CTCTATTGTT	900
82	ACCTCTGACT	CCCAGTTGTT	TAATAAACCA	TATTGGTTAC	ATAAGGCACA	GGGTCATAAC	960
83	AATGGTATCT	GCTGGCATAA	TCAATTATTT	GTTACTGTGG	TAGATACCAC	TCGTAGTACC	1020
84	AATTTAACAA	TATGTGCTTC	TACACAGTCT	CCTGTACCTG	GGCAATATGA	TGCTACCAAA	1080
85	TTTAAGCAGT	ATAGCAGACA	TGTTGAAGAA	TATGATTTGC	AGTTTATTTT	TCAGTTATGT	1140
86	ACTATTACTT	TAAGTGCAGA	TGTTATGTCC	TATATTCATA	GTATGAATAG	CAGTATTTTA	1200
87	GAGGATTGGA	ACTTTGGTGT	TCCCCCCCCG	CCAATACTA	GTTTGGTGGA	TACATATCGT	1260
88	TTTGTACAAT	CTGTTGCTAT	TACCTGTCAA	AAGGATGCTG	CACCAGCTGA	AAATAAGGAT	1320
89	CCCTATGATA	AGTTAAAGTT	TTGGAATGTG	GATTTAAAGG	AAAAGTTTTC	TTTGGACTTA	1380
90	GATCAATATC	CCCTTGACG	TAAATTTTTG	GTTTCAGGCTG	GATTGCGTCG	CAAGCCCACC	1440
91	ATAGGCCCTC	GTAAACGTTT	TGCTCCATCT	GCCACTACGT	CTTCTAAACC	TGCCAAGCGT	1500
92	GTGCGTGAC	GTGCCAGGAA	GTAA				1524

93
94 (2) INFORMATION FOR SEQ ID NO:2:
95

96 (i) SEQUENCE CHARACTERISTICS:

97 (A) LENGTH: 507 amino acids

98 (B) TYPE: amino acid

99 (C) STRANDEDNESS: single

RAW SEQUENCE LISTING PATENT APPLICATION US/08/409,122

DATE: 01/29/97
TIME: 13:48:20

INPUT SET: S15202.raw

```

100      (D) TOPOLOGY: linear
101
102      (ii) MOLECULE TYPE: protein
103      (iii) HYPOTHETICAL: NO
104      (iv) ANTI-SENSE: NO
105      (v) FRAGMENT TYPE: N-terminal
106      (vi) ORIGINAL SOURCE:
107
108      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
109
110      Met Ala Leu Trp Arg Pro Ser Asp Asn Thr Val Tyr Leu Pro Pro Pro
111      1      5      10      15
112      Ser Val Ala Arg Val Val Asn Thr Asp Asp Tyr Val Thr Arg Thr Ser
113      20      25      30
114      Ile Phe Tyr His Ala Gly Ser Ser Arg Leu Leu Thr Val Gly Asn Pro
115      35      40      45
116      Tyr Phe Arg Val Pro Ala Gly Gly Gly Asn Lys Gln Asp Ile Pro Lys
117      50      55      60
118      Val Ser Ala Tyr Gln Tyr Arg Val Phe Arg Val Gln Leu Pro Asp Pro
119      65      70      75      80
120      Asn Lys Phe Gly Leu Pro Asp Asn Ser Ile Tyr Asn Pro Glu Thr Gln
121      85      90      95
122      Arg Leu Val Trp Ala Cys Ala Gly Val Glu Ile Gly Arg Gly Gln Pro
123      100      105      110
124      Leu Gly Val Gly Leu Ser Gly His Pro Phe Tyr Asn Lys Leu Asp Asp
125      115      120      125
126      Thr Glu Ser Ser His Ala Ala Thr Ser Asn Val Ser Glu Asp Val Arg
127      130      135      140
128      Asp Asn Val Ser Val Asp Tyr Lys Gln Thr Gln Leu Cys Ile Leu Gly
129      145      150      155      160
130      Cys Ala Pro Ala Ile Gly Glu His Trp Ala Lys Gly Thr Ala Cys Lys
131      165      170      175
132      Ser Arg Pro Leu Ser Gln Gly Asp Cys Pro Pro Leu Glu Leu Lys Asn
133      180      185      190
134      Thr Val Leu Glu Asp Gly Asp Met Val Asp Thr Gly Tyr Gly Ala Met
135      195      200      205
136      Asp Phe Ser Thr Leu Gln Asp Thr Lys Cys Glu Val Pro Leu Asp Ile
137      210      215      220
138      Cys Gln Ser Ile Cys Lys Tyr Pro Asp Tyr Leu Gln Met Ser Ala Asp
139      225      230      235      240
140      Pro Tyr Gly Asp Ser Met Phe Phe Cys Leu Arg Arg Glu Gln Leu Phe
141      245      250      255
142      Ala Arg His Phe Trp Asn Arg Ala Gly Thr Met Gly Asp Thr Val Pro
143      260      265      270
144      Gln Ser Leu Tyr Ile Lys Gly Thr Gly Met Arg Ala Ser Pro Gly Ser
145      275      280      285
146      Cys Val Tyr Ser Pro Ser Pro Ser Gly Ser Ile Val Thr Ser Asp Ser
147      290      295      300
148      Gln Leu Phe Asn Lys Pro Tyr Trp Leu His Lys Ala Gln Gly His Asn
149      305      310      315      320
150      Asn Gly Ile Cys Trp His Asn Gln Leu Phe Val Thr Val Val Asp Thr
151      325      330      335
152      Thr Arg Ser Thr Asn Leu Thr Ile Cys Ala Ser Thr Gln Ser Pro Val

```

RAW SEQUENCE LISTING PATENT APPLICATION US/08/409,122

DATE: 01/29/97

TIME: 13:48:22

INPUT SET: S15202.raw

```

153          340          345          350
154 Pro Gly Gln Tyr Asp Ala Thr Lys Phe Lys Gln Tyr Ser Arg His Val
155          355          360          365
156 Glu Glu Tyr Asp Leu Gln Phe Ile Phe Gln Leu Cys Thr Ile Thr Leu
157          370          375          380
158 Thr Ala Asp Val Met Ser Tyr Ile His Ser Met Asn Ser Ser Ile Leu
159          385          390          395          400
160 Glu Asp Trp Asn Phe Gly Val Pro Pro Pro Pro Thr Thr Ser Leu Val
161          405          410          415
162 Asp Thr Tyr Arg Phe Val Gln Ser Val Ala Ile Thr Cys Gln Lys Asp
163          420          425          430
164 Ala Ala Pro Ala Glu Asn Lys Asp Pro Tyr Asp Lys Leu Lys Phe Trp
165          435          440          445
166 Asn Val Asp Leu Lys Glu Lys Phe Ser Leu Asp Leu Asp Gln Tyr Pro
167          450          455          460
168 Leu Gly Arg Lys Phe Leu Val Gln Ala Gly Leu Arg Arg Lys Pro Thr
169          465          470          475          480
170 Ile Gly Pro Arg Lys Arg Ser Ala Pro Ser Ala Thr Thr Ser Ser Lys
171          485          490          495
172 Pro Ala Lys Arg Val Arg Val Arg Ala Arg Lys
173          500          505
174

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1389 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

191 ATGGTATCCC ACCGTGCCGC ACGACGCAAA CGGGCTTCGG TGA CTGACTT ATATAAAACA 60
192 TGTAACAAT CTGGTACATG TCCATCTGAT GTTGTTAATA AGGTAGAGGG CACCACGTTA 120
193 GCAGATAAAA TATTGCAATG GTCAAGCCTT GGTATATTTT TGGGTGGACT TGGCATAGGT 180
194 ACTGGAAGTG GTACAGGGGG TCGTACAGGG TACATTCCAT TGGGTGGGCG TTCCAATACA 240
195 GTTGTGGATG TCGGTCCTAC ACGTCCTCCA GTGGTTATTG AACCTGTGGG CCCCACAGAC 300
196 CCATCTATTG TTACATTAAT AGAGGACTCA AGTGTTGTTA CATCAGGTGC ACCTAGGCCCT 360
197 ACTTTTACTG GCACGTCTGG GTTTGATATA ACATCTGCTG GTACAACTAC ACCTGCAGTT 420
198 TTGGATATCA CACCTTCGTC TACCTCTGTT TCTATTTCCA CAACCAATTT TACCAATCCT 480
199 GCATTTTCTG ATCCGTCCAT TATTGAAGTT CCACAACTG GGGAGGTGTC AGGTAATGTA 540
200 TTTGTTGGTA CCCCTACATC TGGAACACAT GGGTATGAAG AAATACCTTT ACAAACATTT 600
201 GCTTCTTCTG GTACGGGGGA GGAACCCATT AGTAGTACCC CATTGCCTAC TGTGCGGCGT 660
202 GTAGCAGGTC CCCGCCTTTA CAGTAGGGCC TACCAACAAG TGTCTGTGGC TAACCCTGAG 720
203 TTTCTTACAC GTCCATCCTC TTAAATTACC TATGACAACC CGGCCTTTGA GCCTGTGGAC 780
204 ACTACATTAA CATTTGAGCC TCGTAGTAAT GTTCCTGATT CAGATTTTAT GGATATTATC 840
205 CGTTTACATA GGCCTGCTTT AACATCCAGG CGTGGTACTG TGCGCTTTAG TAGATTAGGT 900

```

RAW SEQUENCE LISTING PATENT APPLICATION US/08/409,122

DATE: 01/29/97

TIME: 13:48:24

INPUT SET: S15202.raw

206 CAAAGGGCAA CTATGTTTAC CCGTAGCGGT ACACAAATAG GTGCTAGGGT TCACTTTTAT 960
 207 CATGATATAA GTCCTATTGC ACCCTCCCCA GAATATATTG AACTGCAGCC TTTAGTATCT 1020
 208 GCCACGGAGG ACAATGGCTT GTTTGATATA TATGCAGATG ACATAGACCC TGCAATGCCT 1080
 209 GTACCATCGC GTCCTACTAC CTCCTCTGCA GTTTCTACAT ATTCGCCCCAC TATATCATCT 1140
 210 GCCTCTTCCT ATAGTAATGT AACGGTCCCT TTAACCTCCT CTTGGGATGT GCCTGTATAC 1200
 211 ACGGGTCCTG ATATTACATT ACCACCTACT ACCTCTGTAT GGCCCATTGT ATCACCACACA 1260
 212 GCCCCTGCCT CTACACAGTA TATTGGTATA CATGGTACAC ATTATTATTT GTGGCCATTA 1320
 213 TATTATTTTA TTCCTAAAAA GCGTAAACGT GTTCCCTATT TTTTTCGAGA TGGCTTTGTG 1380
 214 GCGGCCTAG 1389

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 461 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

231
 232 Met Val Ser His Arg Ala Ala Arg Arg Lys Arg Ala Ser Val Thr Asp
 233 1 5 10 15
 234 Leu Tyr Lys Thr Cys Lys Gln Ser Gly Thr Cys Pro Ser Asp Val Val
 235 20 25 30
 236 Asn Lys Val Glu Gly Thr Thr Leu Ala Asp Lys Ile Leu Gln Trp Ser
 237 35 40 45
 238 Ser Leu Gly Ile Phe Leu Gly Gly Leu Gly Ile Gly Thr Gly Ser Gly
 239 50 55 60
 240 Thr Gly Gly Arg Thr Gly Tyr Ile Pro Leu Gly Gly Arg Ser Asn Thr
 241 65 70 75 80
 242 Val Val Asp Val Gly Pro Thr Arg Pro Pro Val Val Ile Glu Pro Val
 243 85 90 95
 244 Gly Pro Thr Asp Pro Ser Ile Val Thr Leu Ile Glu Asp Ser Ser Val
 245 100 105 110
 246 Val Thr Ser Gly Ala Pro Arg Pro Thr Phe Thr Gly Thr Ser Gly Phe
 247 115 120 125
 248 Asp Ile Thr Ser Ala Gly Thr Thr Pro Ala Val Leu Asp Ile Thr
 249 130 135 140
 250 Pro Ser Ser Thr Ser Val Ser Ile Ser Thr Thr Asn Phe Thr Asn Pro
 251 145 150 155 160
 252 Ala Phe Ser Asp Pro Ser Ile Ile Glu Val Pro Gln Thr Gly Glu Val
 253 165 170 175
 254 Ser Gly Asn Val Phe Val Gly Thr Pro Thr Ser Gly Thr His Gly Tyr
 255 180 185 190
 256 Glu Glu Ile Pro Leu Gln Thr Phe Ala Ser Ser Gly Thr Gly Glu Glu
 257 195 200 205
 258 Pro Ile Ser Ser Thr Pro Leu Pro Thr Val Arg Arg Val Ala Gly Pro

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION: US/08/409,122

DATE: 01/29/97
TIME: 13:48:28

INPUT SET: S15202.raw

Line	Error	Original Text
30	Wrong application Serial Number	(A) APPLICATION NUMBER: 08/408,669